Pattern Recognition for Brain-Computer Interfaces by Combining Support Vector Machine with Adaptive Genetic Algorithm

Banghua Yang, Shiwei Ma, and Zhihua Li

Shanghai Key Laboratory of Power Station Automation Technology, Department of Automation, Shanghai University, Shanghai 200072, China
yangbanghua@shu.edu.cn, swma@mail.shu.edu.cn, lzh_sh@staff.shu.edu.cn

Abstract. Aiming at the recognition problem of EEG signals in brain-computer interfaces (BCIs), we present a pattern recognition method. The method combines an adaptive genetic algorithm (GA) with the support vector machine (SVM). It integrates the following three key techniques: (1) the feature selection and model parameters of the SVM are optimized synchronously, which constitutes a hybrid optimization; (2) the aim of the hybrid optimization is to improve the classification performance of the SVM; and (3) the hybrid optimization is solved by using the adaptive GA. The method is used to classify three types of EEG signals produced during motor imaginations. It yields 72% classification accuracy, which is higher 8% than the one obtained with the individual optimization of the feature selection and SVM parameters.

1 Introduction

A brain-computer interface (BCI) is an alternative communication and control channel that does not depend on the brain's normal output pathway of peripheral nerves and muscles [1]. A BCI system can help severely disabled people to communicate with computers or control electronic devices through their thoughts. Most BCIs utilize EEG signals to detect distinguishable brain states. These distinguishable brain states are then transformed into external actions through the recognition of EEG signals. Over the past years many evidences have evaluated the possibility to recognize a few mental tasks from EEG signals [2-4]. However, how to improve the recognition performance of EEG signals is still a key problem [5].

The recognition procedure of EEG signals includes three steps: the feature extraction, the feature selection and the classification. This paper mainly concerns the feature selection and the classification. The feature selection is to select an optimal feature subset from all candidate features, which is an optimization problem. The feature selection can improve the generalization performance of the classifier, reduce its complexity and speed up its training process. In the classification, parameters of a classifier affect its classification performance. The selection of parameters of the classifier is also an optimization problem. In previous methods used in BCIs, none of these two optimization problems are considered or they are performed independently

K. Li et al. (Eds.): LSMS 2007, LNBI 4689, pp. 307–316, 2007.

[6-10]. However, the feature selection and the classification are dependent on each other. No optimization or optimizing only one of them is difficult to ensure that these two problems obtain optimal solutions simultaneously.

We will explore a novel method that optimizes the feature selection and classifier parameters simultaneously. The support vector machine (SVM) is a relatively new classification technique that has shown to perform strongly in a number of real-world problems, including BCIs [5]. We will use the SVM as the classifier. At the same time, the genetic algorithm (GA) is a global and probabilistic search algorithm that is based on the mechanics of natural selection and population genetics. It can maintain a good balance between searching width and searching deepness [11]. So, we will use an adaptive GA to optimize the Feature Selection and SVM parameters simultaneously, so the method is called GA-FS-SVM.

2 Data

Six healthy subjects (sub1-sub6) participated the experiment. They seated in a shielded room with dim lighting. Sounds around the surroundings were not controlled painstakingly considering for further application. A 32-channel elastic electrode cap was used to record EEG. The data were recorded at the sampling rate 100Hz with ESI-128.

Each subject repeated the experiment for two sessions. Each session comprised 150 trials. The subjects were asked to imagine performing one of three motor imagery tasks (playing basketball with left hand, playing basketball with right hand, and braking with right foot) in a self-paced mode during each trial. Each trial lasted 5.75s~6.25s (mean 6s) and consisted of three phases: 1) a 0.75s~1.25s (random) resting phase; 2) a 1s preparation phase; and 3) a 4s of motor imagery task phase during which subjects were performing the corresponding motor imagery task according to the direction of the arrow (a left arrow and a right arrow indicate to imagine left hand and right hand movement respectively, a down arrow means right foot). The data during the last 4s of each trial were used to perform analysis. The module of the data acquisition can be seen in Fig.1.

3 Method

3.1 The Feature Extraction

Fig.2 depicts the diagram of a simple BCI system. The proposed GA-FS-SVM concerns the feature selection and the classification. As for the feature extraction, we adopt the spectral power as the feature, which is commonly used in BCIs. The most related frequency information during motor imagery is the Mu (8-12Hz) and Beta (18-26Hz) rhythms on the scalp just above the motor cortex [12]. Mean powers within these two bands are calculated as features. So, two dimensional features can be obtained for each EEG channel. Considering the practicality of BCIs system, we use six electrodes (C3, C4, P3, P4, CZ, and PZ, see Fig.3.) which are considered important EEG channels. Then we can obtain a 12-dimensional feature vector $F = \{f_1, f_2, ..., f_{12}\}$, in which $f_1 \sim f_6$ are mean powers within the band 8-12Hz

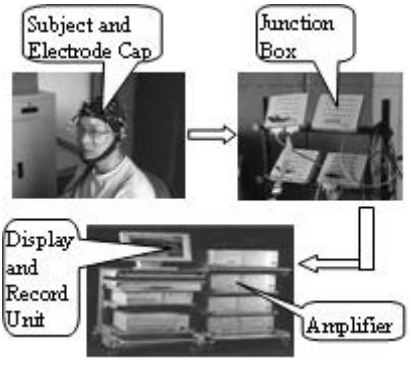

Fig. 1. The module of the data acquisition

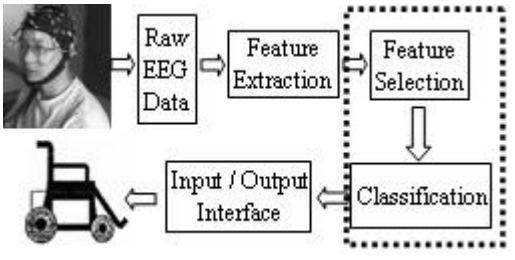

Fig. 2. The diagram of a simple BCI system

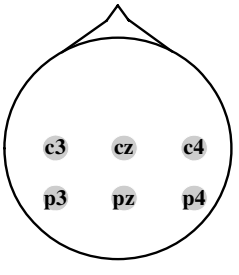

Fig. 3. The used six electrodes

of six channels respectively according to the above channel order and $f_7 \sim f_{12}$ are mean powers within the band 18-26Hz.

3.2 The Adaptive GA Theory

The GA is a global and probabilistic search algorithm that is based on the mechanics of natural selection and population genetics. It can maintain a good balance between searching width and searching deepness. The GA starts to search from a set of initial solutions in a population. An individual (al-so called a chromosome) implies a possible solution to the problem and it consists of many genes. Each gene represents a feature or a parameter. In the feature selection, a binary gene represents a feature in

which a gene bit "1" denotes the corresponding feature is selected and a gene bit "0" denotes the feature is eliminated. The method of optimizing parameters is similar to the feature selection and the only difference between them is that a floating-point gene represents a parameter.

The GA consists of many parameters, such as selection operator, crossover operator, mutation operator, etc. The crossover operator is crossed by probability P_c and the mutation operator is mutated by probability P_m. Fixed crossover probability P_c and mutation probability P_m may result in premature and local convergence. So, we adopt an adaptive GA. The adaptive GA can be defined by the following formulas:

$$P_c = k_1(f_{\max} - f')/(f_{\max} - \overline{f}) \tag{1}$$

$$P_m = k_2(f_{\max} - f'')/(f_{\max} - \overline{f}) \tag{2}$$

Where k_1, k_2 are constants, $k_1, k_2 \le 1.0$, the two parameters should be adjusted according to a given problem, $f_{\max}$ and $\overline{f}$ are the maximum fitness and the average fitness of a population respectively; f' is the larger one of fitness values of the two individuals used to cross; f'' is the fitness of the individual used to mutate. The detailed description of the adaptive GA can be seen in [13].

3.3 The Basic SVM Theory

The SVM is a powerful and relatively new classification method based on statistical learning theory. The SVM has many remarkable characteristics such as good generalization performance, the absence of local minima and sparse representation of the solution. The problem of pattern recognition may be stated as follows: given a data set L, with x_i input features, y_i classification output and m the number of samples, then

$$L = \{(x_1, y_1), (x_2, y_2), ..., (x_m, y_m)\} \tag{3}$$

The SVM finds an optimal separating hyperplane by maximizing the margin between classes. The algorithm consists of solving the following optimization problem:

$$\min_{w,\xi} \frac{1}{2}(w \cdot w) + C(\sum_{i=1}^{l} \xi_i) \tag{4}$$

where $\xi_i \ge 0$, $y_i(w\phi(x_i) + b) \ge 1 - \xi_i$.

The parameter ξ_i is called a slack variable and ensure that the problem has a solution in case the data are not linear separable. The parameter C is a tradeoff variable, w is an adjustable weight vector, $\varphi(x)$ is a nonlinear function for feature mapping. The decision function is

$$f(x) = \sum_{i=1}^{m} w\phi(x_i) + b \tag{5}$$

It can be described further by the dot product,

$$f(x) = \sum_{i=1}^{m} \alpha_i y_i (\phi(x_i) \cdot \phi(x)) + b \tag{6}$$

The dot product can be performed by a Kernel function $K(x, y)$. So, the decision function can be described as follows:

$$f(x) = \sum_{i=1}^{l} \alpha_i y_i K(x_i, x) + b \tag{7}$$

During solving the SVM, finding good Kernel function parameters and a parameter C is an important part of the model selection. The classification performance of the SVM is strongly dependent on values of parameters.

3.4 The GA-FS-SVM Method

The feature vector is $F = \{f_1, f_2, \cdots, f_{12}\}$. We can encode a chromosome with $S = \{s_1, s_2, \cdots, s_{12}\}$, $s_i \in \{0,1\}, i = 1,2,\cdots,12$. Before classification with the SVM, some SVM parameters need to be given. The most common Kernel functions are polynomial function and radial basis function. We select the polynomial function $(Gamma * u \cdot v + Coeff)^{Degree}$, where u, v are input vectors, $Gamma, Coeff, Degree$ are parameters of the kernel function. So the training model of the SVM can be constructed as $M = \{Gamma, Coeff, Degree, C\}$, where $Gamma, Coeff, Degree, C \geq 0$. We can encode a chromosome with $C = \{c_1, c_2, c_3, c_4\}$, where $c_i \in R$, i=1,2,3,4.

The GA-FS-SVM method optimizes the feature selection and SVM parameters simultaneously and its structural diagram is shown in Fig.4. The hybrid optimization can be regarded as optimizing $H = \{F, M\}$. The chromosome H is encoded with $G = \{s_1, s_2, \cdots s_{12}, c_1, c_2, c_3, c_4\}$. A specified chromosome leads to a feature subset and a SVM model simultaneously.

We evaluate the performance (fitness) of a chromosome using the average classification accuracy of the SVM. The calculation of the fitness can be outlined as follows: (1) to a specified chromosome, we randomly select half of all trials as training samples; (2) the selection procedure of training samples is performed ten times and so we can obtain ten values of the classification accuracy; (3) the fitness is obtained by averaging the ten values. When the adaptive GA attains convergence, the optimal chromosome is obtained, i.e. the optimal feature subset and SVM parameters are obtained. The optimized results are used to classify unknown samples.

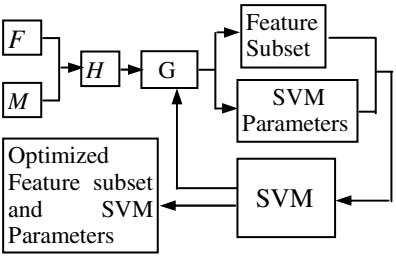

Fig. 4. The structural diagram of the GA-FS-SVM

3.5 Parameters of the Adaptive GA

(1) *Parameters initialization*: We select the evolution generation t=100, population size p=100. The selected ranges of $Gamma, Coeff, Degree, C$ are [0,2], [0,5], [0,1], [0,500] respectively according to our experience.

(2) *Selection*: we adopt the selection mechanism of proportional fitness and elitism strategy. The chance of reproduction for an individual of the parent generation to the next generation is proportional to its fitness value. Meanwhile, the fittest individual is taken over directly into the next generation.

(3) *Crossover*: we select a single-point crossover mechanism with a probability of p_c =0.8 to create new chromosomes in each pair. The crossover probability p_c is obtained according to formula (1).

(4) *Mutation*: we adopt a multi-uniform mutation operator combining with a multi-Gaussian mutation operator, in which the mutation probability of each operator is 0.05. The mutation probability of each operator p_m is obtained according to formula (2). In formula (1) and (2), k_1=0.8; 6) k_2=0.4. k_1 and k_2 .are obtained by the experience and adjustments. The other parameters are selected according to common suggestions in [11].

4 Results and Analysis

4.1 Results

In order to verify the performance of the GA-FS-SVM, we perform the following three strategies:

(1) *GA-FS-SVM*: it is described in section 3;

(2) *GA-FS*: the adaptive GA only optimizes the feature selection. As for the classification, we randomly select five groups of SVM parameters: C1{0.1, 1, 0.1, 300}, C2{0.2, 0.1, 0.1, 200}, C3{1, 0.1, 0.1 300}, C4{1.5, 0.1, 0.1, 250}, C5{1.2, 3, 0.1, 400}.

(3) *GA-SVM*: the adaptive GA only optimizes SVM parameters. As for the feature subset, we randomly select five groups: F1 {f_1, f_2, f_5}, F2 {f_2, f_5, f_8, f_9, f_{11}}, F3 {f_3, f_5, f_7, f_9}, F4 {f_5, f_6, f_7, f_9, f_{10}}, F5 {$f_1 \sim f_{12}$}.

It should be noted that initial features, GA parameters and the calculation of the fitness used in all above methods are the same. As an example, Fig.5 shows the classification accuracy of training samples of subject 1 (sub1) versus generation t with different strategies. Classification results of testing samples of different subjects with different strategies are shown in Tab.1. It also should be noted that results of the GA-FS-SVM are obtained by running the program ten times and then averaging the ten values.

We calculate the mean value and the standard deviation of classification accuracies. It should be noted that these values are obtained by calculating C1~C5 in the GA-FS, F1~F5 in the GA-SVM. In addition, we perform the one sample t test to the GA-FS-SVM with other strategies. The mean value, the standard deviation, and the t value are shown in Tab.2. The comparison of mean values of the classification accuracy among different strategies is plotted in Fig.6.

4.2 Analysis

From Tab.1 we can see that different strategies can result in different results for any one subject. In addition, different parameters in one strategy can also result in different classification accuracies, which show that the feature selection and SVM parameters all affect the classification performance.

The Fig.6 shows that the GA-FS-SVM obtains the best result among all strategies because it optimizes the feature selection and the classification synchronously and so it obtains the optimal feature subset and SVM parameters synchronously. Results of t test in Tab.2 show that the classification accuracy obtained by the GA-FS-SVM is significantly higher than other methods. The GA-FS only optimizes the feature selection and the GA-SVM only optimizes SVM parameters, which means that they lack optimal SVM parameters or the optimal feature subset. So obtained results by them are inferior to the one obtained by the GA-FS-SVM. The GA-FS-SVM obtains an average classification accuracy (mean value of six subjects) 72.0%, which is higher about 8% than the one (64.2%, by averaging six subjects and the GA-FS, the GA-SVM) obtained by the GA-FS and the GA-SVM.

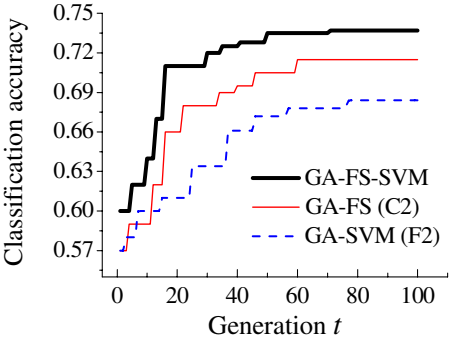

Fig. 5. The classification accuracy versus generation t

Table 1. Classification results of different subjects with different strategies

Method name		GA-FS -SVM	GA-FS					GA-SVM				
			C1	C2	C3	C4	C5	F1	F2	F3	F4	F5
Classification accuracy (%)	sub1	70.6	67.2	65.4	67.5	68.6	66.5	64.7	62.9	63.6	62.8	61.7
	sub2	72.4	65.6	63.8	64.2	67.5	64.8	66.1	64.8	65.1	64.1	62.8
	sub3	68.5	61.8	59.7	60.1	61.6	59.8	62.9	64.1	63.1	63.7	65.1
	sub4	74.1	64.3	62.9	61.7	63.6	60.5	68.5	69.1	66.8	65.7	66.2
	sub5	68.9	63.5	60.7	61.5	59.8	60.5	63.7	65.8	61.9	62.7	65.3
	sub6	75.8	65.8	64.7	62.9	63.1	64.8	69.1	68.7	66.8	67.4	69.1

Table 2. The mean value, the standard deviation and t value

subjects		sub1	Sub2	Sub3	Sub4	Sub5	Sub6
Mean value (%)	GA-FS	67.0	65.2	60.6	62.6	61.2	64.3
	GA-SVM	63.1	64.6	63.8	67.3	63.9	68.2
Standard deviation (%)	GA-FS	1.19	1.46	1.02	1.52	1.42	1.23
	GA-SVM	1.11	1.23	0.88	1.47	1.66	1.06
t (9)	p<0.01	-7.501	-18.12	-10.50	-10.24	-9.91	-12.86

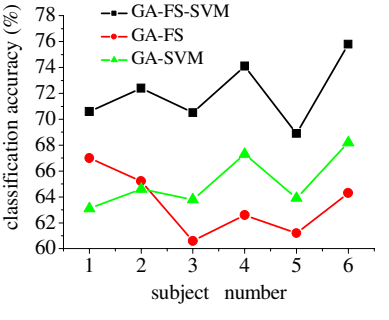

Fig. 6. The comparison of different strategies

5 Conclusions

(1) Both the feature selection and SVM parameters play an important role in the classification. Different feature subset and different SVM parameters can result in different classification results.
(2) In BCIs, the GA-FS-SVM optimizes the feature selection and the SVM parameters synchronously, which can pick the most promising feature subset and excellent training model to classification. It avoids the disadvantage of optimizing only one of them.

Duo to the limited amount of data and subjects, the classification accuracy needs to be further investigated. Based on the very promising results we obtained here, we are investigating the possibility of developing the GA-FS-SVM further. In this paper, we want to show the potential of the hybrid optimization method.

Acknowledgements. The research is partly supported by Shanghai Leading Academic Discipline Project (T0103).

References

1. Wolpaw, J.R., Birbaumer, N., Heetderks, W.J.: Brain computer interface technology: a review of the first international meeting. IEEE Trans. Rehab. Eng. 8, 64–73 (2000)
2. Millán, J.d.R., Mouriño, J., Franzé, M., Cincotti, F.: A local neural classifier for the recognition of EEG patterns associated to mental tasks. IEEE Trans. Neural Networks 13, 678–686 (2002)
3. Millán, J.d.R.: Adaptive brain interfaces. Commun. ACM 46, 74–80 (2003)
4. Wolpaw, J.R., McFarland, D.J., Neat, G.W., Forneris, C.A.: An EEG-based brain-computer interface for cursor control. Electroenceph. Clin. Neurophysiol. 78, 252–259 (1991)
5. schroder, M., Bogdan, M., Rosenstiel, W., Hinterberger, T. (ed.): Automated EEG Feature Selection for Brain Computer Interfaces. In: Proc. 1st Int. IEEE EMBS Conf. Neural Engineering, Capri Island, Italy, pp. 626–629 (2003)
6. Garrett, D., Peterson, D.A., Anderson, C.W., Thaut, M.H.: Comparison of Linear, Nonlinear, and Feature Selection Methods for EEG Signal Classification. IEEE Trans. Neural Syst. Rehab. Eng. 11, 141–144 (2003)
7. Graimann, B., Huggins, J.E., Levine, S.P., Pfurtscheller, G.: Detection of ERP and ERD/ERS patterns in single ECoG channels. In: Proc. 1st Int. IEEE EMBS Conf. Neural Engineering, Capri Island, Italy, pp. 614–617 (2003)
8. Yom-Tov, E., Inbar, G.F.: Feature selection for the classification of movements from single movement-related potentials. IEEE Trans. Neural Syst. Rehab. Eng. 10, 170–177 (2002)
9. Kaper, M., Meinicke, P., Grossekathoefer, U., Lingner, T., Ritter, H.: BCI Competition 2003—Data Set IIb: Support Vector Machines for the P300 Speller Paradigm. IEEE Trans. Biomed. Eng. 51, 1073–1076 (2004)
10. Lal, T.N., Schröder, M., Hinterberger, T., Weston, J., Bogdan, M., Birbaumer, N., Schölkopf, B.: Support vector channel selection in BCI. IEEE Trans. Biomed. Eng. 51, 1003–1010 (2004)

11. Guangnan, X., Ruiwei, C.: Genetic Algorithm and Engineering Optimization. Tsinghua, Beijing, China (2004)
12. Brett, D.M., Justin, W., Sebastian Seung, H.: BCI competition 2003-Data Set Ia: combining gamma-band power with slow cortical potentials to improve single-trial classification of electroencephalographic signals. IEEE Trans. Biomed. Eng. 51, 1052–1056 (2004)
13. Guozheng, Y., Ting, W., Banghua, Y.: Automated feature selection based on an adaptive genetic algorithm for brain-computer interfaces. LNCS, pp. 575–582 (2006)